Driver drowsiness monitoring system based on facial Landmark detection with convolutional neural network for prediction

Roaa Albasrawi[1], Fajer F. Fadhil[2], Mohammed Talal Ghazal[1]
[1]Department of Computer Engineering Technology, Northern Technical University, Mosul, Iraq
[2]Department of Computer Engineering, Mosul University, Mosul, Iraq

Article Info

Article history:

Received Apr 16, 2022
Revised Jun 20, 2022
Accepted Jul 28, 2022

Keywords:

Advanced driver assistance systems
Convolutional neural network
Drowsiness
EfficientNet
Prediction

ABSTRACT

Several factors often contribute to car accidents, most of them caused by human error, and the most notable are drowsiness, fatigue, distracted driving, and alcohol. Although self-driving cars are the best solution to save human lives and avoid car accidents, they are expensive. The roads in many countries are not prepared for the movement of this type of car. Scare new technologies included in modern cars, such as backup cameras and sensors, contributed to keeping drivers safer in this paper. A driver monitoring system is based on determining the driver's face's main points, which provide the required vital information for face analysis. The EfficientNet convolutional neural network (ConvNet) model is used for facial landmarks prediction, which is employed to detect face drowsiness and fatigue in real-time. The system is trained to detect multiple traits, including facial expressions, yawning and head poses. The results show that employing facial landmarks will assist in efficiently producing eyes and mouth features, which can assist in appropriately creating models to analyze drowsiness. Due to this, the proposed safety features are applicable and available in future vehicles.

Corresponding Author:

Mohammed Talal Ghazal
Department of Computer Engineering Technology, Northern Technical University
Mosul 41002, Iraq
Email: mohammed.ghazal@ntu.edu.iq

1. INTRODUCTION

According to the National Highway Transportation Safety Administration (NHTSA) studies in 2016, 2018, and 2019 about 96% of all car accidents are caused by human mistakes [1]-[3]. These studies indicate many factors contributing to vehicle accidents, such as drowsiness and fatigue caused by lack of sleep, driving drunk, and looking down at the phone to read or send a text message. Vehicle accidents can be avoided with advanced driver assistance systems (ADAS) [4]. The ADAS can support the driver by the vehicle surrounding environment information to prevent accidents. The major ADAS safety applications include automatically emergency braking, recognition of traffic signs, lane departure alert, pedestrian avoidance, and blind spot detection [5]. These applications represent the key to lifesaving by utilizing the latest sensor systems and running vision-based algorithms. Our goal is to design a reliable driver safety system with the lowest cost to help in the modern automobile design process. Our work in this paper focuses on the vehicle interior, especially on a driver's facial trait. Yawning and facial expressions are among the most prominent traits that help in driver fatigue detection [6]-[8]. It is commonly known that yawning represents another sign of drowsiness; head drooping indicates fatigue; moreover, anger, fear, surprise, and sadness negatively affect the driver [9], [10]. The system will be more robust if it integrates all these decision-making factors. The deep learning approach save been used to observe a ddriver'sbehavior based on face monitoring. Driving behavior is several actions of a driver while going. It can be classified into normal and abnormal driving, where everyday driving is defined as the typical daily behavior. In contrast, the weird, such as drowsy and drunk driving, is uncommon and results from some physical or mental factors [11]. Deep learning is a form of machine learning that makes the computer learn through experiments to collect knowledge and understand the world [12], [13]. One of the deep learning classes is a convolutional neural network (CNN), commonly utilized for image analysis [14]-[16]. It mimics the neurons connectivity method in human brains [17], [18]. Figure 1 clarifies the CNN structure; it consists of three essential parts: input images, feature extraction, and classification. The CNN needs less preprocessing for input images than other classification algorithms for the feature extraction stage; it's made up of multiple layers, containing convolutional layers, rectified linear unit (ReLU) activation function pooling layers, and normalization layers. There are fully connected layers for the classification stage and one classification layer [19], [20].

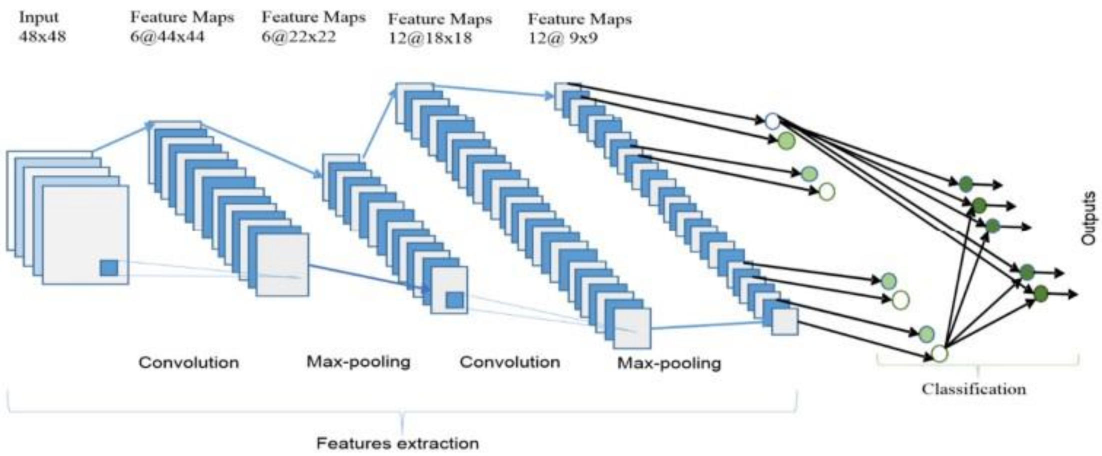

Figure 1. The CNN architecture [13]

One of the main reasons for the profound learning superiority of traditional image processing methods is the way of representing the input data. Moreover, the performance increases with good data representation [21]-[23]. Thus, many researchers focused on feature extraction from raw data based on machine learning methods [24]. Traditional methods such as scale-invariant feature transform (SIFT), histogram of oriented gradients (HOG), and speeded up robust features (SURF) need a lot of time and effort in feature extraction. At the same time, the deep learning algorithms can perform it automatically [25], [26]. This work aims to detect the facial landmark through a particular group of points, such as the corners of (the eye, mouth, or nose). Monitoring the information of face can help understand the driver's situation and prevent fatal accidents. This paper's main contribution is to use the EfficientNet CNN model for facial landmarks prediction, which is mapped to recognize the shape of the human face.

2. DROWSINESS DETECTION METHODS

There are many drowsiness and fatigue detection methods based on driver behavior; these methods depend on several parameters to figure out the driver's fatigue [27]. Yawning, eye blinking, and facial features are among these parameters. Various conventional methods to detect the driver's drowsiness have been summarized, followed by the newest deep learning methods.

2.1. The conventional methods:

- Eye blink monitoring method: Rahman *et al*. [28], an eye blinking strategy has been proposed to detect drowsiness. Harris corner algorithm is used to detect eyes corners on both sides. The eye state of open/closed is determined based on calculating the distance from the center to the bottom of an eye within a known time interval.
- Yawning detection method: Abtahi *et al*. [29], the driver's tiredness identification strategy is based on estimating his physiological, social, and execution state. To recognize tiredness, they focused on yawn identification with three strategies: recognizing face and mouth, recognizing face dependent on layout coordination, then mouth using shading condition, and finally utilizing Viola-Jones hypothesis for recognizing face and mouth to identify yawn.

- Wearable electroencephalographic (EEG) system method: Li *et al*. [30], the driver's sleepiness has been recognized based on EEG signal. In this work, 20 subjects (15 for training/and 5 for the test) are utilized for mind exercises to identify the driver's laziness. They proposed three levels of drowsiness, starting with an alert, then early warning, and drowsy. Support vector machine is used to examine the level of tiredness, and the system gives the precision of drowsy detection (91.92%), early-warning (83.78%), and (91.25%) for alert.
- A hybrid strategy: Oliveira *et al*. [31], a real-road experiment has been utilized to compare the performance of electrooculogram (EOG) and electrocardiogram (ECG) methods to detect driver drowsiness. The hybrid strategy by combining both ways increased the system robustness rather than using the EOG method Individually.
- Multiple sensors method: Anand *et al*. [32], an Arduino-based system with various sensors has been proposed to detect driver drowsiness, including an eye blink sensor, an alcohol sensor, and a heartbeat sensor. The system monitors the sensor's status; for any abnormal state, the vehicle automatically slows down and stops.
- Facial features monitoring method: Manu [33], face detection and skin segmentation have been proposed to detect driver weariness. The edge detection algorithm is used for eye tracking. The K-means algorithm is utilized for yawning detection, and system accuracy was 94.58%.
- In addition to these researches, many studies and tests are conducted to detect driver fatigue and drowsiness. Despite some studies that obtained good results, these methods have computational and applied complexities; moreover, they cannot satisfy the real-time requirements.

2.2. The deep learning methods

The superiority of deep learning over traditional methods of solving complex problems has led to its being widely used. Deep learning is extensively utilized for computer vision purposes like object detection, emotion recognition, and image classification. There are several approaches to detecting drowsiness using deep understanding. For instance, Vijayan and Sherly [34], a combination of three deep learning models to extract facial features has been proposed to compose a feature-fused architecture (FFA). Experimentally ResNet50, VGG16, and InceptionV3 model of CNN has been used; regardless of the three networks and FFA, the model of InceptionV3 has shown an accuracy rate of 78%. Park *et al*. [35], researchers followed the same approach for drowsiness detection by combining the obtained results of AlexNet, VGG-FaceNet, and FlowNet with fully connected layers. The proposed architecture is called deep drowsiness detection (DDD), which achieved 73.06% of drowsiness detection accuracy. In one more similar approach, in [36], the researcher used a hybrid of CNN and long short-term memory (LSTM) for drowsiness detection.

3. METHOD

Figure 2 shows the proposed system, which consists of three steps—starting with the detection of the driver's face, while the second is the facial landmarks prediction based on the EfficientNet model. The final step is drowsiness detection by calculating the number of eyes blinking and yawning. Whenever drivers feel drowsy, the rate of eye closure frequency raises. If this rate overrides the threshold value, then the system must generate an alarm. The multi-task cascaded convolutional networks (MTCNN) [37] are used for face detection because it's one of the most accurate and fast face detectors. The CNN model based on the EfficientNet network is used for landmark prediction. The Google mind group has developed the baseline of the EfficientNet network; they proposed a more efficient model as suggested by its name. The EfficientNet model architecture is unlike the traditional convolutional network design that primarily concentrates on selecting the appropriate layer architecture; EfficientNet employs the concept of compound scaling to expand the model size (length, width, and image resolution) without modifying the predefined architecture in the baseline model to enhance the model accuracy [38] as shown in Figure 3(a) and Figure 3(b).

In the compound scaling method, a compound coefficient ($\emptyset$) is used for scaling network dimensions: width (w), depth (d), and resolution (r) in a principled way:

$$
\begin{aligned}
& d = \alpha^{\emptyset} \\
& w = \beta^{\emptyset} \\
& r = \gamma^{\emptyset} \\
& s.t\ \alpha.\beta^2.y^2 \approx 2, \qquad (1) \\
& \alpha \geq 1, \beta \geq 1, \gamma \geq 1
\end{aligned}
$$

the dimensions (α, β *and* γ) are constants that a small grid search can determine.

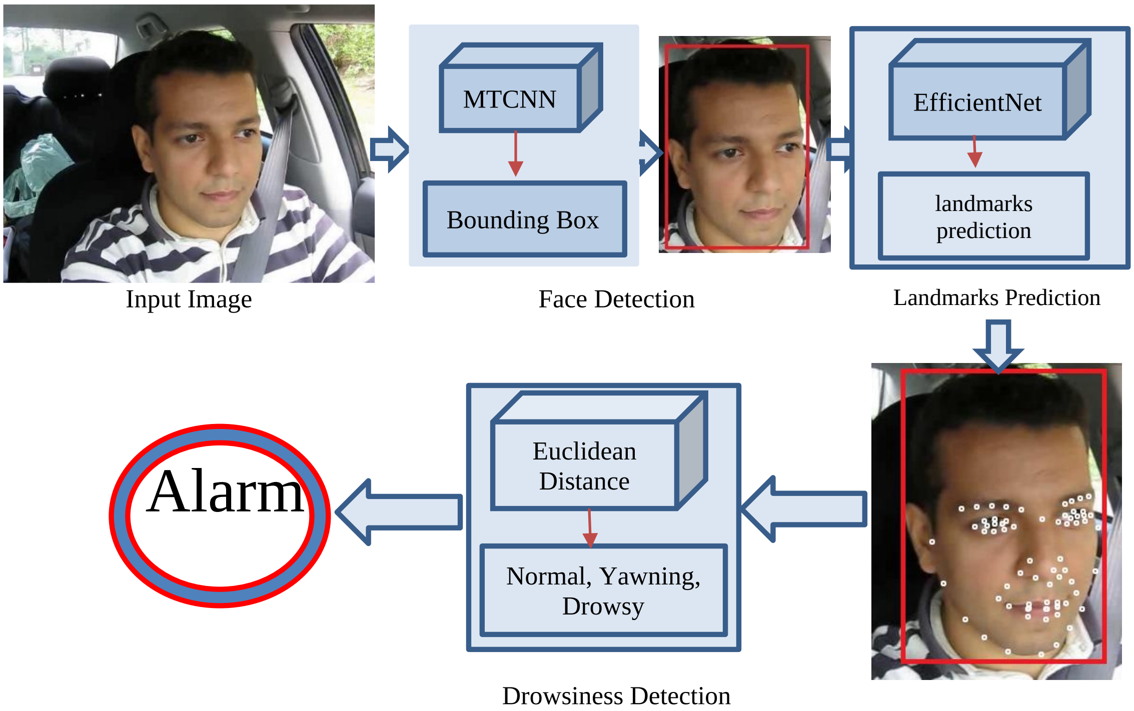

Figure 2. The proposed system block

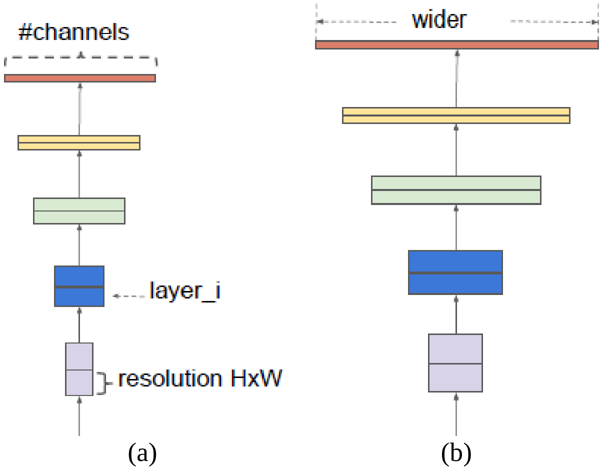

Figure 3. Model scaling (a) is a baseline of the conventional network and (b) is a conventional scaling of network dimension

We can conclude from equation that the coefficient can control the available resources to expand the model. In contrast, the depth, width, and resolution can hold how to manage these additional network resources. The predicted facial landmarks from the EfficientNet model are used to locate the coordinates of facial features. The coordinates shown in Figure 4 are employed to track the eyes and mouth distance ratio.

We consider the eyes and mouth contours to determine the drowsy based on facial contours (1). The eye aspect ratio (EAR) is used to compute the distance between the eyes and from edge to edge based on the euclidean distance formula. In the same way, the distance between the mouth lips edges is calculated to predict the rate of yawn frequency times.

$$EAR = \frac{\|P2-P6\|+\|P3-P5\|}{2\|P1-P4\|} \tag{2}$$

Physically, when the driver gets drowsy, the eyelids become closer while the lips getaway. This increases the distance between the lips while decreasing the space between the eyelids in a yawning state. The frequency of (f) for both eye and yawn are calculated within a specific time to determine the drowsy threshold (θ). If $(f > parent\ \theta)$ t, the alarm will alert the driver by displaying a frame message.

Figure 4. The coordinates of facial landmarks

4. EXPERIMENTAL WORK AND RESULTS

The 300 W face dataset is used for facial landmarks prediction and other images as well. The dataset considers the diversity of expression, identity, illumination conditions, head poses, and face size. The images were semi-automatically annotated with the 68-point mark-up. The model has trained on NVIDIA's 2nd gen RTX 3060 graphics processing unit (GPU) architecture and with a Tensorflow environment. The images are resized to (128×128), and the dataset has been split into 80% for training stage and 20% for testing. The model was trained using 100 batch sizes and 100 epochs. Adopmizethree with 10^{-3} learning rate and categorical crossentropy loss function has been used in the experimental work to predict the facial landmarks with the EfficientNet model. The model accuracy is approximately 82%, with a 0.5 dropout rate, as shown in Figure 5. The model is robust in detecting facial landmarks, especially the area around the eyes and nose. Finally, these landmarks have been mapped to calculate the aspect ratio of the eyes and the mouth, to achieve yawn detection and closed eye recognition. To achieve the best result, many videos have been taken to test the precision of the drowsiness detection system, using different kinds of cameras with different resolutions. Figure 6 shows a video taken of a driver with a closed eye and opened mouth as a test result.

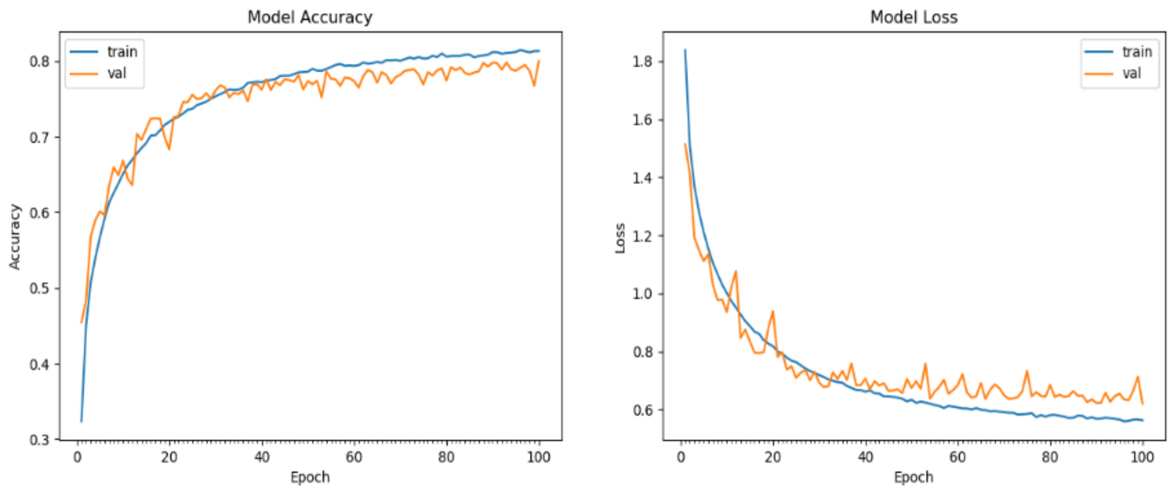

Figure 5. EfficientNet model accuracy and loss

Figure 6. Drowsy detection results

5. CONCLUSION

This work considered the recognition of eye and mouth status for helping to judge drowsiness. A model of ConvNet based on EfficientNet has been proposed to predict the facial landmarks, which are mapped to predict the facial key points on the input face in real-time. From the above results, it is clear that the proposed model is effective, reliable, and usable. In the future, the proposed model can be implemented on Android applications to avoid accidents caused by driver drowsiness.

ACKNOWLEDGEMENTS

This research is funded by cooperation between Northern Technical University and Mosul University in Iraq. Website: for Northern Technical University https://www.ntu.edu.iq and Website: for Mosul University https://www.uomosul.edu.iq.

REFERENCES

[1] National Highway Traffic Safety Administration, "2016 motor vehicle crashes: overview," *Traffic safety facts: research note*, Oct. 2017.
[2] National Highway Traffic Safety Administration, "2018 motor vehicle crashes: overview," *Traffic safety facts: research note*, Oct. 2019.
[3] National Highway Traffic Safety Administration, "Preview of motor vehicle traffic fatalities in 2019," *Research Note, DOT HS*, vol. 813, p, 021, Oct. 2020.
[4] A. Ziebinski, R. Cupek, D. Grzechca, and L. Chruszczyk, "Review of advanced driver assistance systems (ADAS)," in *AIP Conference Proceedings*, vol. 1906, no. 1, p. 120002, 2017, doi: 10.1063/1.5012394.
[5] M. M. Antony and R. Whenish, "Advanced driver assistance systems (ADAS)," in *Automotive Embedded Systems: Springer*, 2021, pp. 165-181, doi: /10.1007/978-3-030-59897-6_9.
[6] M. Omidyeganeh *et al*., "Yawning detection using embedded smart cameras," in *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 3, pp. 570-582, Mar. 2016, doi: 10.1109/TIM.2015.2507378.
[7] M. Jeong and B. C. Ko, "Driver's facial expression recognition in real-time for safe driving," *Sensors*, vol. 18, no. 12, p. 4270, Dec. 2018, doi: 10.3390/s18124270.
[8] N. D. Al-Shakarchy and I. Ali, "Detecting abnormal movement of driver's head based on spatial-temporal features of video using deep neural network DNN," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 19, no. 1, pp. 344-352, Jul. 2020, doi: 10.11591/ijeecs.v19.i1.pp344-352.
[9] N. H. Ali, A. R. Abdullah, N. M. Saad, A. S. Muda, T. Sutikno, and M. H. Jopri, "Brain stroke computed tomography images analysis using image processing: A review," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 10, no. 4, pp. 1048-1059, Dec. 2021, doi: 10.11591/ijai.v10.i4.pp1048-1059.
[10] E. J. Baker *et al*., "User identification system for inked fingerprint pattern based on central moments," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 24, no. 2, pp. 1149-1160, Nov. 2021, doi: 10.11591/ijeecs.v24.i2.pp1149-1160.
[11] S. Al-Sultan, A. H. Al-Bayatti, and H. Zedan, "Context-aware driver behavior detection system in intelligent transportation systems," in *IEEE Transactions on Vehicular Technology*, vol. 62, no. 9, pp. 4264-4275, 2013, doi: 10.1109/TVT.2013.2263400.
[12] K. G. Kim, "Book review: deep learning," *Healthcare informatics research*, vol. 22, no. 4, pp. 351-354, Oct. 2016, doi: 10.4258/hir.2016.22.4.351.
[13] N. Y. Abdullah, M. T. Ghazal, and N. Waisi, "Pedestrian age estimation based on deep learning," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 22, no. 3, pp. 1548-1555, Jun. 2021, doi: 10.11591/ijeecs.v22.i3.pp1548-1555.
[14] N. Waisi, N. Y. Abdullah, and M. Ghazal, "The automatic detection of underage troopers from live-videos based on deep learning," *Przegląd Elektrotechniczny*, vol. 2021, no. 9, pp. 85-88, Sep. 2021, doi: 10.15199/48.2021.09.18.
[15] M. Berrahal and M. Azizi, "Augmented binary multi-labeled CNN for practical facial attribute classification," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 23, no. 2, pp. 973-979, Aug. 2021, doi: 10.11591/ijeecs.v23.i2.pp973-979.

[16] S. Hosgurmath, V. V. Mallappa, N. B. Patil, and V. Petli, "A face recognition system using convolutional feature extraction with linear collaborative discriminant regression classification," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 2, pp. 1468-1476, Aug. 2022, doi: 10.11591/ijece.v12i2.pp1468-1476.

[17] F. H. K. Zaman, "Gender classification using custom convolutional neural networks architecture," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 6, pp. 5758-5771, Dec. 2020, doi: 10.11591/ijece.v10i6.pp5758-5771.

[18] M. Z. Alom *et al*., "A state-of-the-art survey on deep learning theory and architectures," *Electronics*, vol. 8, no. 3, p. 292, Mar. 2019, doi: 10.3390/electronics8030292.

[19] M. Karthikeyan and T. Subashini, "Automated object detection of mechanical fasteners using faster region based convolutional neural networks," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 6, pp. 430-5437, Dec. 2021, doi: 10.11591/ijece.v11i6.pp5430-5437.

[20] I. Salehin *et al*., "Analysis of student sentiment during video class with multi-layer deep learning approach," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 4, pp. 3981-3993, Aug. 2022, doi: 10.11591/ijece.v12i4.pp3981-3993.

[21] S. Pouyanfar *et al*., "A survey on deep learning: algorithms, techniques, and applications," *ACM Computing Surveys (CSUR)*, vol. 51, no. 5, pp. 1-36, Sep. 2019, doi: 10.1145/3234150.

[22] I. Hadi and A. Mahdi, "Generating images of partial face using landmark based k-nearest neighbor," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 17, no. 1, pp. 420-428, Jan. 2020, doi: 10.11591/ijeecs.v17.i1.pp420-428.

[23] M. A. Naji, G. A. Salman, and J. F. Muthna, "Face recognition using selected topographical features," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 5, pp. 4695-4700, Oct. 2020, doi: 10.11591/ijece.v10i5.pp4695-4700.

[24] M. T. Ghazal and K. J. T. Abdullah, "Face recognition basedon curvelets, invariant moments features and SVM," *TELKOMNIKA Telecommunication, Computing, Electronics and Control*, vol. 18, no. 2, pp. 733-739, Apr. 2020, doi: 10.12928/TELKOMNIKA.v18i2.14106.

[25] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, ‘‘Deep learning applications and challenges in big data analytics,’’ *Journal of Big Data*, vol. 2, no. 1, pp. 1-21, 2015, doi: 10.1186/s40537-014-0007-7.

[26] M. Ghazal, N. Waisi, and N. J. T. Abdullah, "The detection of handguns from live-video in real-time based on deep learning," *TELKOMNIKA Telecommunication, Computing, Electronics and Control*, vol. 18, no. 6, pp. 3026-3032, Dec. 2020, doi: 10.12928/telkomnika.v18i6.16174.

[27] M. Kumar, "Driver drowsiness detection techniques: a review," *Advances in Mathematics: Scientific Journal*, vol. 9, no. 6, pp. 3933–3938, Jun. 2020, doi: 10.37418/amsj.9.6.73.

[28] A. Rahman, M. Sirshar, and A. Khan, "Real time drowsiness detection using eye blink monitoring," *2015 National Software Engineering Conference (NSEC)*, 2015, pp. 1-7, doi: 10.1109/NSEC.2015.7396336.

[29] S. Abtahi, B. Hariri, and S. Shirmohammadi, "Driver drowsiness monitoring based on yawning detection," *2011 IEEE International Instrumentation and Measurement Technology Conference*, 2011, pp. 1-4, doi: 10.1109/IMTC.2011.5944101.

[30] G. Li, B. -L. Lee, and W. -Y. Chung, "Smartwatch-based wearable EEG system for driver drowsiness detection," in *IEEE Sensors Journal*, vol. 15, no. 12, pp. 7169-7180, Dec. 2015, doi: 10.1109/JSEN.2015.2473679.

[31] L. Oliveira, J. S. Cardoso, A. Lourenço, and C. Ahlström, "Driver drowsiness detection: a comparison between intrusive and non-intrusive signal acquisition methods," *2018 7th European Workshop on Visual Information Processing (EUVIP)*, 2018, pp. 1-6, doi: 10.1109/EUVIP.2018.8611704.

[32] S. J. Anand, Mahendaren, M. A. Kumar, K. Pugalarasu, and K. Surya, " Drowsy and drunken drive control, automatic accident detection and rescue system using Arduino" *International Journal of Scientific Development and Research (IJSDR)*, vol. 5, no. 12, pp. 23-29, Dec. 2020.

[33] B. N. Manu, "Facial features monitoring for real time drowsiness detection," *2016 12th International Conference on Innovations in Information Technology (IIT)*, 2016, pp. 1-4, doi: 10.1109/INNOVATIONS.2016.7880030.

[34] V. Vijayan and E. Sherly, “Real time detection system of driver drowsiness based on representation learning using deep neural networks,” *Journal of Intelligent & Fuzzy Systems*, vol. 36, no. 3, pp. 1977–1985, 2019, doi: 10.3233/JIFS-169909.

[35] S. Park, F. Pan, S. Kang, and C. D. Yoo, “Driver drowsiness detection system based on feature representation learning using various deep networks,” *Asian Conference on Computer Vision*, Springer, Cham, vol. 10118, pp. 154–164, Nov. 2016, doi: 10.1007/978-3-319-54526-4_12.

[36] J.-M. Guo and H. Markoni, “Driver drowsiness detection using hybrid convolutional neural network and long short-term memory,” *Multimedia tools and applications*, vol. 78, no. 20, pp. 29059–29087, Jul. 2018, doi: 10.1007/s11042-018-6378-6.

[37] M. Ghazal, R. Albasrawi, N. Waisi, and M. Al Hammoshi, "Smart meeting attendance checking based on a multi-biometric recognition system," *Przegląd Elektrotechniczny*, vol. 2022, no. 3, pp. 93-96, Mar. 2022, doi: 10.15199/48.2022.03.21.

[38] M. Tan and Q. Le, "Efficientnet: rethinking model scaling for convolutional neural networks," *International conference on machine learning, PMLR*, 2019, pp. 6105-6114.

BIOGRAPHIES OF AUTHORS

Roaa Albasrawi received the “C.E” Bachelor of Computer Engineering degree and “M.E.” Master of Computer Engineering degree from the Technical College of Mosul, Mousl, Iraq in 2005, 2012 respectively. Her master thesis title is “Iris recognition using circular wavelet filters.” Currently she is a lecturer at Technical Engineering College of Mousl, Mosul, Iraq. Her current research interest includes deep learning, image processing, and cloud computing. She can be contacted at email: roaa.albasrawi@ntu.edu.iq.

Fajer F. Fadhil received the "C.E" Bachelor of Computer Engineering degree and "M.E." Master of Computer Engineering in Electrical and Computer Engineering from Technical College of Mosul, Mosul, Iraq, in 2003 and 2012 respectively. She is currently working at Computer and Internet center at University of Mosul, Mosul, Iraq as a lecturer. She was working as Cisco Academy administrator at University of Mosul, Mosul, Iraq. Her current research interest includes optical engineering, communication engineering and cloud computing. She can be contacted at email: fajrfehr@uomosul.edu.iq.

Mohammed Talal Ghazal obtained his M.Sc. degree from Computer Engineering Technology, Northern Technical University, Mosul, Iraq in 2016. His M.Sc. thesis entitled: "Wheelchair Robot Control Using EOG signals" and his current research focuses on the development of face recognition algorithm. He can be contacted at email: mohammed.ghazal@ntu.edu.iq.